Final scientific report

Overall summary and statistics

Cover Page

BARD Project Number: IS-4062-07

Date of Submission of the report: 1/12/2010

Project Title: Dissection of whitefly-geminivirus interactions at the transcriptomic, proteomic and cellular levels

Investigators	**Institutions**
Principal Investigator (PI):	
Dr. Murad Ghanim	ARO
Co-Principal Investigator (Co-PI):	
Dr. Joe Cicero	U of Arizona
Collaborating Investigators:	
Dr. Judith K. Brown	U of Arizona
Dr. Henryk Czosnek	Hebrew University

Keywords *not* appearing in the title and in order of importance. Avoid abbreviations.
Genomics, genes expression, virus transmission, Tomato, localizarion

Abbreviations commonly used in the report, in alphabetical order**:**

FISH, qRT-PCR, SLCV, TYLCV

Budget: IS: $157,000 US: $118,000 Total: $275,000

Signature
Principal Investigator

Signature
Authorizing Official, Principal Institution

Final scientific report

Publication Summary (numbers)

	Joint IS/US authorship	US Authors only	Israeli Authors only	Total
Refereed (published, in press, accepted) BARD support acknowledged		2	6	8
Submitted, in review, in preparation		2	1	3
Invited review papers			1	1
Book chapters		1 (in prep)		1
Books				
Master theses			1	1
Ph.D. theses			1	1
Abstracts	1	6	3	10
Not refereed (proceedings, reports, etc.)				

Postdoctoral Training: List the names and social security/identity numbers of all postdocs who received more than 50% of their funding by the grant.

Cooperation Summary (numbers)

	From US to Israel	From Israel to US	Together, elsewhere	Total
Short Visits & Meetings		2	1	3
Longer Visits (Sabbaticals)				

Description Cooperation:

Patent Summary (numbers)

	Israeli inventor only	US inventor only	Joint IS/US inventors	Total
Submitted				
Issued (allowed)				
Licensed				

Abstract

Our project focuses on gene expression and proteomics of the whitefly *Bemisia tabaci* (Gennadius) species complex in relation to the internal anatomy and localization of expressed genes and virions in the whitefly vector, which poses a major constraint to vegetable and fiber production in Israel and the USA. While many biological parameters are known for begomovirus transmission, nothing is known about vector proteins involved in the specific interactions between begomoviruses and their whitefly vectors. Identifying such proteins is expected to lead to the design of novel control methods that interfere with whitefly-mediated begomovirus transmission. The project objectives were to: 1) Perform gene expression analyses using microarrays to study the response of whiteflies (B, Q and A biotypes) to the acquisition of begomoviruses (*Tomato yellow leaf curl* (TYLCV) and *Squash leaf curl* (SLCV). 2) Construct a whitefly proteome from whole whiteflies and dissected organs after begomovirus acquisition. 3) Validate gene expression by q-RTPCR and sub-cellular localization of candidate ESTs identified in microarray and proteomic analyses. 4) Verify functionality of candidate ESTs using an RNAi approach, and to link these datasets to overall functional whitefly anatomical studies. During the first and second years biological experiments with TYLCV and SLCV acquisition and transmission were completed to verify the suitable parameters for sample collection for microarray experiments. The parameters were generally found to be similar to previously published results by our groups and others. Samples from whole whiteflies and midguts of the B, A and Q biotypes that acquired TYLCV and SLCV were collected in both the US and Israel and hybridized to *B. tabaci* microarray. The data we analyzed, candidate genes that respond to both viruses in the three tested biotypes were identified and their expression that included quantitative real-time PCR and co-localization was verified for HSP70 by the Israeli group. In addition, experiments were undertaken to employ *in situ* hybridization to localize several candidate genes (in progress) using an oligonucleotide probe to the primary endosymbiont as a positive control. A proteome and corresponding transcriptome to enable more effective protein identification of adult whiteflies was constructed by the US group. Further validation of the transmission route of begomoviruses, mainly SLCV and the involvement of the digestive and salivary systems was investigated (Cicero and Brown). Due to time and budget constraints the RNAi-mediated silencing objective to verify gene function was not accomplished as anticipated. HSP70, a strong candidate protein that showed over-expression after TYLCV and SLCV acquisition and retention by *B. tabaci,* and co-localization with TYLCV in the midgut, was further studies. Besides this protein, our joint research resulted in the identification of many intriguing candidate genes and proteins that will be followed up by additional experiments during our future research. To identify these proteins it was necessary to increase the number and breadth of whitefly ESTs substantially and so whitefly cDNAs from various libraries made during the project were sequenced (Sanger, 454). As a result, the proteome annotation (ID) was far more successful than in the initial attempt to identify proteins using Uniprot or translated insect ESTs from public databases. The extent of homology shared by insects in different orders was surprisingly low, underscoring the imperative need for genome and transcriptome sequencing of homopteran insects. Having increased the number of EST from the original usable 5500 generated several years ago to >600,000 (this project+NCBI data mining), we have identified about one fifth of the whitefly proteome using these new resources. Also we have created a database that links all identified whitefly proteins to the PAVEdb-ESTs in the database, resulting in a useful dataset to which additional ESTS will be added. We are optimistic about the prospect of linking the proteome ID results to the transcriptome database to enable our own and other labs the opportunity to functionally annotate not only genes and proteins involved in our area of interest (whitefly mediated transmission) but for the plethora of other functionalities that will emerge from mining and functionally annotating other key genes and gene families in whitefly metabolism, development, among others. This joint grant has resulted in the identification of numerous candidate proteins involved in begomovirus transmission by *B. tabaci.* A next major step will be to capitalize on validated genes/proteins to develop approaches to interfere with the virus transmission.

Achievements

Our main goal was to discover genes and proteins involved in the circulative transmission of begomoviruses. To do so, we proposed to study the transcriptomic, proteomic and cellular response of whiteflies after acquisition of the begomoviruses, TYLCV and SLCV. We have accomplished most of our stated goals and brief details are provided herein.

Transcriptomics-Israeli Team: whitefly RNA samples were collected in the USA and Israel. The US partner provided RNA from midguts of whiteflies from the A and B biotypes that had acquired SLCV while RNA samples from midgets, heads and whole insects that had acquired TYLCV were collected by the Israeli team (See Appendix –Table1 1 for selected genes with putative relevance to virus-vector interactions). One protein was of special interest and was followed by experiments resulting in a MSc thesis. The *hsp70* gene (RNA) was expressed in whiteflies that had acquired either TYLCV or SLCV (Fig. 1A). Over-expression of HSP70 occurred in whiteflies that retained TYLCV (Fig.1B), and *in vitro* binding between TYLCV and HSP70 was demonstrated (Figure 1C). Co-localization of HSP70 and TYLCV was observed in midgut epithelial cells and physical interactions were confirmed by FISH- and immune-localization (Fig.1D). Other candidate genes selected also expected to mediate virus transmission.

Proteome/Transcriptome-UA Team: The main objectives were to (1) increase the number and breadth of annotated whitefly ESTs and to use them and available non-whitefly insect ESTs (Genbank) to (2) identify a partial whitefly proteome (~385 proteins). The 774,065 ESTs (cDNAs) were obtained at different times and from different sources. The EST sequences were assembled and resulted in 15,045 contigs and 39,864 singletons for a total of 54,909 unique transcripts. The average length of a contig was ~800 bases, with 55% of the contigs containing less than 500 bp and 92% of the contigs containing less than 1000 bp (See Appendices 3,4).

Construction of proteome search libraries from whitefly and other insect ESTs. PAVE was used for assembly. ESTs included Sanger sequences from both a normalized WF library and a WF gut-specific library (this project), and 454 sequences from whole and gut-specific cDNA libraries (this project). Unique transcripts (UniTrans) were translated in six frames using' transeq' (EMBL) and the UA supercomputer. The resultant file was used to search the LC/MS/MS spectra.

Selected Insect EST library. ESTs from *Acyrthosiphon pisum, Aedes aegypti, Anopheles gambiae, Apis mellifera, Bombyx mori, Culex pipiens, Culex quinquefasciatus, Myzus persicae, Nasonia vitripennis, Toxoptera citricida* and *Tribolium castaneum* were downloaded from the NCBI EST. Aphid proteins and contig data from *A. gossypii*, *M. persicae*, and *T. citricida* were downloaded from aphidbase.org. We used standalone blast against the Insecta protein database created above and perl scripts to select ESTs which were longer than 50 basepairs

(or amino acids) and had less than 80% identity using blastx (for EST) or blastp (for Aphid proteins) (See Appendix 2). They were then translated in three forward frames using transeq (EMBL-EBI) (477,017 records) and used to search the LC/MS/MS spectra.

Criteria for protein search. Scaffold (version Scaffold_2_05_01, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they exceeded specific database search engine thresholds. Sequest identifications required at least deltaCn scores of greater than 0.08 and XCorr scores of greater than 1.8, 2.5, 3.5 for singly, doubly, triply charged peptides. X! Tandem identifications required at least -Log(Expect scores) scores of >3.0. Protein identifications were accepted only if they had 2+ peptides. Proteins containing similar peptides that could not be differentiated (MS/MS analysis) alone were grouped by principles of parsimony.

Peptide and Protein Identification. Spectra were searched using Sequest and X!Tandem programs. The resultant dta files were loaded into Scaffold (Version 2.6, Proteome Software, Portland, OR). To assess if a spectrum was identified by more than one theoretical peptide in the search libraries, the spectrum report for each library was exported from Scaffold. MS-Excel was used to identify spectra which were used in multiply-charged states for identification (e.g. the same spectrum was identified in a +2 and a + 3 state). For these spectra (approximately one fifth of the total spectra identified with 90% peptide identification confidence), the peptide match criteria were made more stringent. When these criteria did not eliminate one of the identifications, the identification with the higher delta cn was used.

To obtain GO function (unique identifier or GI number of the top BLASTX hit) each whitefly consensus was fed into the Batch retrieval (PIR) at the UniProt Website *http://www.pir.uniprot.org*. The results were outputted to produce a tabular format that included the Protein ID/accession, protein name, length, organism, taxon/group, UniRef90/50, matched fields, COG ID, Pfam name, gene name, KEGG pathway ID, EC number, GO ID, GO term, e-value and GO slim categories.

Whitefly proteins were identified using the available whitefly ESTs and a 'customized' set of insect EST and protein sequences (NCBI), previously and up to October 2010 using BLASTN and BLASTX to search against NCBI nr/nt), proteins, unigenes, and ESTs for 12 insect genera or species, including the recently published pea aphid genome. GO-Slim functions were assigned at http://www.geneontology.org and protein functions were assigned using Kegg to identify a subset of candidate proteins possibly relevant to whitefly-mediated begomovirus transmission (See Appendix 4).

Possibly informative whitefly genes identified were Actin, Cyclophilin, GAPDH, GBLP (homolog of receptor for activated kinase1, RACK 1), Knottin, salivary gland proteins, membrane-bound proteins. Actin has

been implicated in *Beet western yellow virus* transmission, and associates with the HIV Nef protein. Cyclophilin, an ER protein, is implicated in the transmission of *Cereal yellow dwarf virus-RP*, and in the parasitic protozoan *Toxoplasma gondii*, where it competitively binds to receptors, preventing HIV-1 binding. GroEL homolog (chaperonin 60), a heat shock protein involved in stress, defense, and protein folding, interacts with TYLCV coat protein *in vitro* and is involved in *Potato leaf roll virus* transmission by aphids. Hsp60 interacts with the Human hepatitis B virus polymerase. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is involved in *Beet western yellows virus* transmission in aphids, and in signal transduction during stress-induced apoptosis, and GADPH expression increases during Vaccinia infection. Guanine nucleotide like binding protein (GBLP), a RACK1 homolog and G-protein, is involved in signaling and regulation of cell receptors involved in aphid-mediated transmission of *BWYV*, and interacts with mumps virus V-protein.

Candidate genes were validated for presence in the whitefly genome and for expression by analyzing the transcriptome using PCR and RT-PCR respectively, followed by DNA sequencing of cloned PCR and RT-PCR products (data not shown). Except got knottin, both the gene and transcript were present in whitefly DNA and RNA extracts, respectively. The PCR fragments and transcripts were cloned and sequenced for each candidate gene (3 clones each). Semi-quantitative or 'real-time' polymerase chain reaction (qPCR) primers were designed and optimized to facilitate the amplification and quantification of six genes and their respective transcripts (mRNA) among a suite of whitefly genes of interest.

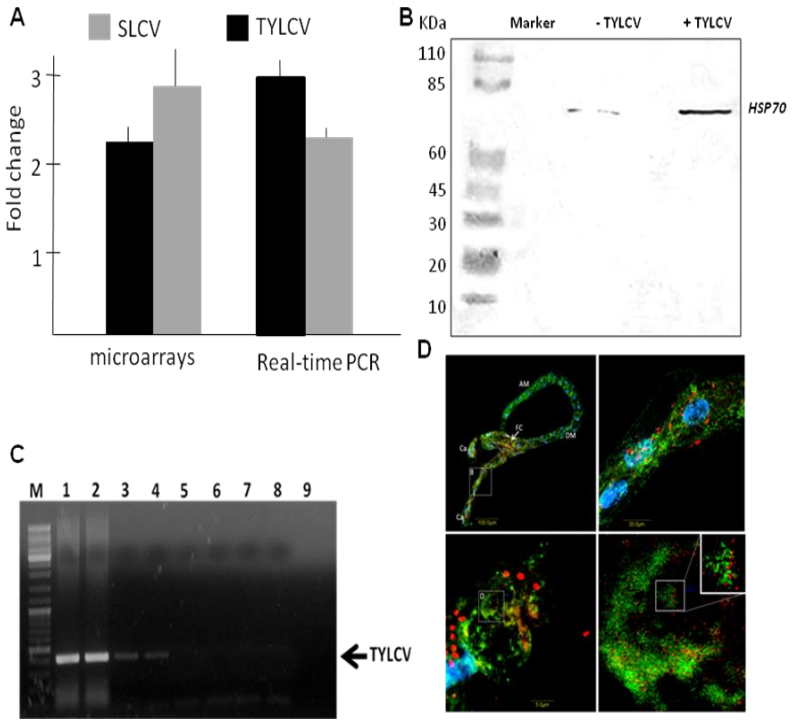

Figure 1. Overexpression of *hsp70* in microarray and real-time PCR analyses after TYLCV retention in *B. tabaci* (A). over expression was also observed at the protein level (B). *in vivo* binding between TYLCV and HSP70 was observed after immunocapture-PCR analysis (C). Co-localization and physical interaction between HSP70 (green) and TYCV (red) was observed in midgut epithelial cells after combined FISH and immunolocalization analyses (D).

Details of Cooperation: Coordination of sample collection for microarrays and proteome construction started from the first day of grant activation in both Israel (TYLCV) and Arizona (SLCV). Samples for RNA construction were sent from the US to Israel and were used in the microarray hybridizations. E-mail was the main communication route between the groups, and e-mails were often circulated and exchanged between the members of both groups when decisions regarding any step of grant implementation were needed. Dr. Ghanim visited the laboratory of Dr. Brown in The University of Arizona during December 2007 and March 2009, and discussed with Drs. Cicero, Brown and S. Hunter the different aspects of the grant, experimental design and expected contributions of both groups during the course of the grant. A meeting between Drs. Brown and Czosnek was also held during the 3rd European Whitefly Symposium that was held in Almeria, Spain Oct 2008 where symposia papers also were presented by our team. Coordination to publish additional manuscripts is underway.

***List of Publications (in which BARD is acknowledged)*:**

Journal Articles

Ghanim, M., Brumin, B., Popovski, S. 2009. A simple, rapid and inexpensive method for localization of *Tomato yellow leaf curl virus* and *Potato leafroll virus* in plant and insect vectors. 159:311-4.

Mahadav, A., Gerling, D., Gottlieb, Y., Czosnek, H., Ghanim, M. 2008. Parasitization by the wasp *Eretmocerus mundus* induces transcription of genes related to immune response and symbiotic bacteria proliferation in the whitefly *Bemisia tabaci.* BMC Genomics. 9:342.

Ghanim, M. and Kontsedalov, S. (2009). Susceptibility to insecticides in the Q biotype of *Bemisia tabaci* is correlated with bacterial symbiont densities. *Pest Manag. Sci*. 65:939-42.

Mahadav, A., Kontsedalov, S., Czosnek, H. and Ghanim, M. (2009). Thermotolerance and gene expression following heat stress in the whitefly bemisia tabaci B and Q biotypes. *Insect Biochem. Mol. Biol.* 39:668-76.

Gottlieb, Y., Zchori-Fein, E., Mozes Daube, N., Kontsedalov, S., Skaljac, M., Brumin, M., Sobol, I., Czosnek, H., Vavre, F., Fleury, F. and Ghanim, M. (2010). The transmission efficiency of *Tomato yellow leaf curl virus* by the whitefly *Bemisia tabaci* is correlated with the presence of a specific symbiotic bacterium species. *J. Virol*. 84:9310-9317. **(with journal cover).**

Cicero, J. M. and J. K. Brown. (2011). Functional anatomy of whitefly organs sssociated with *Squash leaf curl virus* (Geminiviridae: *Begomovirus*) transmission by the B Biotype of *Bemisia tabaci* (Aleyrodidae: Hemiptera). Ann. Entomol. Soc. Am. (accepted).

Cicero, J. and Brown, J.K. (2011). The anatomy of the accessory salivary glands of the whitefly *Bemisia tabaci* (Aleyrodidae:Hemiptera) and correlations to begomovirus transmission. Ann. Entomol. Soc. Am. (accepted).

Appendices

Published papers

Ghanim, M., Brumin, B., Popovski, S. 2009. A simple, rapid and inexpensive method for localization of *Tomato yellow leaf curl virus* and *Potato leafroll virus* in plant and insect vectors. 159:311-4.

Mahadav, A., Gerling, D., Gottlieb, Y., Czosnek, H., Ghanim, M. 2008. Parasitization by the wasp *Eretmocerus mundus* induces transcription of genes related to immune response and symbiotic bacteria proliferation in the whitefly *Bemisia tabaci.* BMC Genomics. 9:342.

Ghanim, M. and Kontsedalov, S. (2009). Susceptibility to insecticides in the Q biotype of *Bemisia tabaci* is correlated with bacterial symbiont densities. *Pest Manag. Sci*. 65:939-42.

Mahadav, A., Kontsedalov, S., Czosnek, H. and Ghanim, M. (2009). Thermotolerance and gene expression following heat stress in the whitefly bemisia tabaci B and Q biotypes. *Insect Biochem. Mol. Biol.* 39:668-76.

Gottlieb, Y., Zchori-Fein, E., Mozes Daube, N., Kontsedalov, S., Skaljac, M., Brumin, M., Sobol, I., Czosnek, H., Vavre, F., Fleury, F. and Ghanim, M. (2010). The transmission efficiency of *Tomato yellow leaf curl virus* by the whitefly *Bemisia tabaci* is correlated with the presence of a specific symbiotic bacterium species. *J. Virol*. 84:9310-9317. **(with journal cover)**

One copy of each 'in press', 'accepted' or 'submitted' paper

Cicero, J. M. and J. K. Brown. (2011). Functional anatomy of whitefly organs sssociated with *Squash leaf curl virus* (Geminiviridae: *Begomovirus*) transmission by the B Biotype of *Bemisia tabaci* (Aleyrodidae: Hemiptera). Ann. Entomol. Soc. Am. (accepted).

Cicero, J. and Brown, J.K. (2011). The anatomy of the accessory salivary glands of the whitefly *Bemisia tabaci* (Aleyrodidae:Hemiptera) and correlations to begomovirus transmission. Ann. Entomol. Soc. Am. (accepted).

Unpublished data briefly summarized.

Appendix 1-Table 1: candidate genes responsive to the presence of begomoviruses in the different whitefly biotypes

Appendix 2- Summary of EST annotation

Appendix 3- Selected ESTs

Appendix 4- Proteome summary